

DEVICE, METHOD, AND RECORDING MEDIUM FOR VOICE DATABASE GENERATION

Patent number: JP2001306087
Publication date: 2001-11-02
Inventor: TAKAMI JUNICHI
Applicant: RICOH CO LTD
Classification:
- international: G10L15/04; G10L15/06; G10L13/06
- european:
Application number: JP20000131529 20000426
Priority number(s):

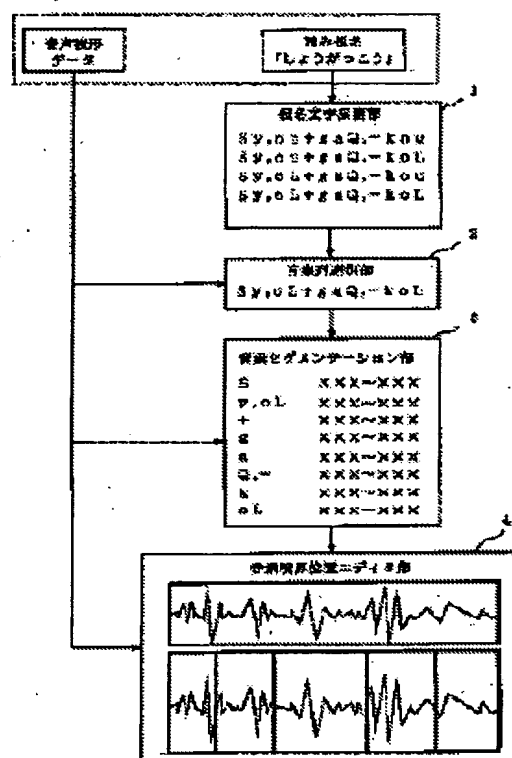
Also published as:

JP2001306087 (

Abstract of JP2001306087

PROBLEM TO BE SOLVED: To greatly reduce the load accompanying voice database generation on an operator.

SOLUTION: This device has a KANA (Japanese syllabary) character expansion part 1, which expands a reading which is described in KANA into phoneme series candidates which can be possibly obtained when the reading is pronounced, a phoneme series selection part 2 which selects the phoneme series bet matching the actual voice data most among the phoneme series candidates expanded by the expansion part 1, a phoneme segmentation part 3 which computes the boarder positions of respective phonemes of voice data in accordance with the phoneme series selected by the selection part 2, and a phoneme border position editor part 4 for interactively correcting the results obtained by the selection part 2 and segmentation part 3.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-306087

(P2001-306087A)

(43) 公開日 平成13年11月2日 (2001.11.2)

(51) Int.Cl.⁷

G 1 0 L 15/04

15/06

13/06

識別記号

F I

G 1 0 L 3/00

5/04

テマコト* (参考)

5 1 5 C 5 D 0 1 5

5 2 1 L

5 2 1 C

E

審査請求 未請求 請求項の数 9 O L (全 9 頁)

(21) 出願番号 特願2000-131529(P2000-131529)

(22) 出願日 平成12年4月26日 (2000.4.26)

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72) 発明者 鷹見 淳一

東京都大田区中馬込1丁目3番6号 株式

会社リコー内

(74) 代理人 100090240

弁理士 植本 雅治

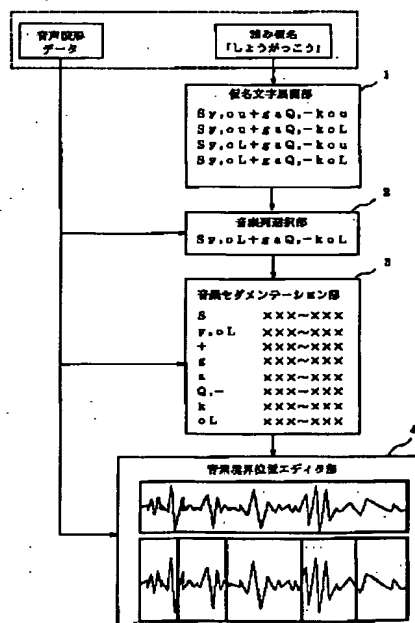
Fターム(参考) 5D015 FF07 HH05 HH11 LL04 LL05

(54) 【発明の名称】 音声データベース作成装置および音声データベース作成方法および記録媒体

(57) 【要約】

【課題】 音声データベース作成に伴う作業者の負担を著しく軽減させる。

【解決手段】 仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開する仮名文字展開部1と、仮名文字展開部1で展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択する音素列選択部2と、音素列選択部2で選択された音素列に従って音声データの各音素の境界位置を算出する音素セグメンテーション部3と、音素列選択部2および音素セグメンテーション部3で得られた結果を対話的に修正するための音素境界位置エディタ部4とを有している。



【特許請求の範囲】

【請求項1】 仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開する仮名文字展開部と、仮名文字展開部で展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択する音素列選択部と、音素列選択部で選択された音素列に従って音声データの各音素の境界位置を算出する音素セグメンテーション部と、音素列選択部および音素セグメンテーション部で得られた結果を対話的に修正するための音素境界位置エディタ部とを有していることを特徴とする音声データベース作成装置。

【請求項2】 請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用いることを特徴とする音声データベース作成装置。

【請求項3】 請求項2記載の音声データベース作成装置において、前記音素セグメンテーション部は、複数の候補から仮名文字展開部で展開された個々の音素の境界位置の平均および分散を求める際に、複数の探索経路から求められる音素の境界位置の情報に対して、その経路のスコアに応じた重みを乗じて集計することを特徴とする音声データベース作成装置。

【請求項4】 請求項2記載の音声データベース作成装置において、前記音素セグメンテーション部は、大量の候補を高速に算出するために、A*探索法を利用することを特徴とする音声データベース作成装置。

【請求項5】 請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音声セグメンテーション部において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値を提示することを特徴とする音声データベース作成装置。

【請求項6】 請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算

出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音素セグメンテーション部において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値に応じて、カーソルの色を変化させることを特徴とする音声データベース作成装置。

【請求項7】 請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音素セグメンテーション部において得られたそれぞれの音素境界位置の分散に応じて、マニュアル操作で移動可能な音素境界位置の範囲に制限を設けることを特徴とする音声データベース作成装置。

【請求項8】 仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開し、展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択させ、選択された音素列に従って音声データの各音素の境界位置を算出し、算出された各音素の境界位置を対話的に修正することで、音声データベースを作成することを特徴とする音声データベース作成方法。

【請求項9】 仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開し、展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択させ、選択された音素列に従って音声データの各音素の境界位置を算出し、算出された各音素の境界位置を対話的に修正することで、音声データベースを作成する処理をコンピュータに実行させるためのプログラムを記録したコンピュータ読取可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声データベース作成装置および音声データベース作成方法および記録媒体に関する。

【従来の技術】高性能な音声認識や、高音質な音声合成を行うためには、音声認識用の高精度な音響モデル、あるいは音声合成用の高品質な音声素片が必要であり、それらの学習、あるいは抽出を行うための音声データベースの整備が不可欠となる。

【0002】音声データベースの作成を行う際に、もっとも厄介な問題は、大量に収集した音声サンプルに対し

て、いかに少ない人的労力で、高い精度の音素ラベル情報を付与するかという点である。

【0003】ここで、音素ラベル情報の付与とは、連続して発声された音声データに対して、音声の波形や周波数スペクトルなどを参考にしながら、それに含まれる個々の音素の種類を記述した音素ラベル、およびその開始時刻と終了時刻に関する情報を付与する作業であり、一般に、その作業にはかなりの労力と熟練が要求される。

【0004】

【発明が解決しようとする課題】本発明は、音声データベース作成に伴う作業者の負担を著しく軽減させることの可能な音声データベース作成装置および音声データベース作成方法および記録媒体を提供することを目的としている。

【0005】

【課題を解決するための手段】上記目的を達成するために、請求項1記載の発明は、仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開する仮名文字展開部と、仮名文字展開部で展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択する音素列選択部と、音素列選択部で選択された音素列に従って音声データの各音素の境界位置を算出する音素セグメンテーション部と、音素列選択部および音素セグメンテーション部で得られた結果を対話的に修正するための音素境界位置エディタ部とを有していることを特徴としている。

【0006】また、請求項2記載の発明は、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用いることを特徴としている。

【0007】また、請求項3記載の発明は、請求項2記載の音声データベース作成装置において、前記音素セグメンテーション部は、複数の候補から仮名文字展開部で展開された個々の音素の境界位置の平均および分散を求める際に、複数の探索経路から求められる音素の境界位置の情報に対して、その経路のスコアに応じた重みを乗じて集計することを特徴としている。

【0008】また、請求項4記載の発明は、請求項2記載の音声データベース作成装置において、前記音素セグメンテーション部は、大量の候補を高速に算出するために、A*探索法を利用することを特徴としている。

【0009】また、請求項5記載の発明は、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モ

デルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音声セグメンテーション部において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値を提示することとを特徴としている。

【0010】また、請求項6記載の発明は、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音素セグメンテーション部において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値に応じて、カーソルの色を変化させることを特徴としている。

【0011】また、請求項7記載の発明は、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音素セグメンテーション部において得られたそれぞれの音素境界位置の分散に応じて、マニュアル操作で移動可能な音素境界位置の範囲に制限を設けることを特徴としている。

【0012】また、請求項8記載の発明は、仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開し、展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択させ、選択された音素列に従って音声データの各音素の境界位置を算出し、算出された各音素の境界位置を対話的に修正することで、音声データベースを作成することを特徴としている。

【0013】また、請求項9記載の発明は、仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開し、展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択させ、選択された音素列に従って音声データの各音素の境界位置

を算出し、算出された各音素の境界位置を対話的に修正することで、音声データベースを作成する処理をコンピュータに実行させるためのプログラムを記録したコンピュータ読取可能な記録媒体を特徴としている。

【0014】

【発明の実施の形態】以下、本発明の実施形態を図面に基づいて説明する。図1は本発明に係る音声データベース作成装置の構成例を示す図である。図1を参照すると、この音声データベース作成装置は、音声認識のための音響モデル学習用サンプルの作成や、音声合成のための音声素片の作成などの用途に使用される音声ラベル付きの音声データベースを作成するためのものであって、仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開する仮名文字展開部1と、仮名文字展開部1で展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択する音素列選択部2と、音素列選択部2で選択された音素列に従って音声データの各音素の境界位置を算出する音素セグメンテーション部3と、音素列選択部2および音素セグメンテーション部3で得られた結果を対話的に修正するための音素境界位置エディタ部4とを有している。

【0015】ここで、仮名文字展開部1は、仮名文字で表記された読み情報から音素記号列への展開を行なう機能を有している。仮名文字で表記された読み情報から音素記号列への展開を行なうための具体的な処理内容は、最終的な音素体系をどのように定めるかに依存するが、一般的には、以下の3段階の変換により実現することができる。

【0016】すなわち、第1段階として、表記記号としての仮名文字から表音記号としての仮名文字への展開を行なう。

【0017】日本語の表記文字としての仮名文字は、ほとんどのものが実際の発音と一対一に対応しているが、エ段母音の後の「い」や、オ段母音の後の「う」に関しては、文字通り「い」や「え」と発音される他に、先行母音の長音化という形で発音される場合がある。例えば「そうさ」を発音する場合、「ソウサ」と「ソーサ」の2通りの可能性が存在する。

【0018】また、が行の音節については、子音“g”の音が鼻音化する場合としない場合の2通りの可能性が存在する。例えば「にほんご」を発音する場合、「ニホンゴ」と「ニホンコ」の2通りの可能性が存在する(ただし「コ」は、鼻音化した「ゴ」の音を表すものとする)。

【0019】このような規則を考慮して、表記記号としての仮名文字を表音記号としての仮名文字へ展開する規則は、一般に一对多対応の変換規則となり、例えば次の

“Q -” → “Q, -”

の2つを融合ラベルとする)

“_ G” → “_ + g”

ように記述することができる。

【0020】「こう」 → 「コウ」、「コー」

「が」 → 「ガ」、「カ」

「っ」 → 「ッ」

【0021】この規則を、表記記号としての仮名文字に対してこの表の出現順に繰返し適用する(展開規則に複数の可能性がある規則を適用する場合には、その数だけ候補の複製を作成した後、それぞれの規則を適用する)ことによって、表音記号としての仮名文字候補を得ることができる。

【0022】例えば、「がっこう」の場合、まず「こう」の部分(この時点で「がっコウ」と「がっコー」の2つの候補が得られる)、次に「が」の部分(この時点で「ガッコウ」と「ガッコー」と「カッコウ」と「カッコー」の4つの候補が得られる)、最後に「っ」の部分がそれぞれ変換されて、最終的に、「ガッコウ」と「ガッコー」と「カッコウ」と「カッコー」の4つの候補が得られることになる。

【0023】なお、この部分の処理で使用される変換規則は、採用する音素体系には依存しない。

【0024】次に、第2段階として、表音記号としての仮名文字から音素列への展開を行なう。

【0025】この段階では、個々の表音文字から実際の音素並びへの変換を行う。母音“i”や“u”の無声化の可能性についても、この段階で考慮する。

【0026】この部分の処理で使用される変換規則は、採用する音素体系に依存するが、この部分も第1段階と同様、一般に一对多対応の変換規則により記述される。

【0027】この変換規則は、例えば次のようになる。

【0028】「カ」 → “G a”

「ガ」 → “+ g a”

「ッ」 → “Q”

「コ」 → “- K o”

「ウ」 → “u”

「ー」 → “L”

【0029】この変換により、「ガッコウ」「ガッコー」「カッコウ」「カッコー」の4通りの表音記号は、“_ + g a Q - k o u _” “_ + g a Q - k o L _” “_ G a Q - k o u _” “_ G a Q - k o L _”の4通りの音素列に展開される。

【0030】次に、第3段階として、音素コンテキストを考慮した音素列の変換を行なう。すなわち、第3段階では、第2段階までの変換で考慮されていない音素コンテキストの影響を反映させるための変換を行なう。

【0031】このための規則は、例えば次のようになる。

(促音と無音の区別はできないため、こ

(語頭のが行音は鼻音化しない)

“L” → “L”

(長音記号は先行母音にくっつける)

【0032】この規則を適用した場合、同じ音素列の候補が複数生成される可能性があるため、そのような候補は1つで代表させることで、最終的な音素列を得ることができる。

【0033】例えば、第2段階で得られた「_ + g a Q - k o u _」「_ + g a Q - k o L _」「_ G a Q - k o u _」「_ G a Q - k o L _」の4通りの音素列については、最終的に「_ + g a Q, - k o u _」「_ + g a Q, - k o L _」の2つの音素列が得られる。

【0034】また、音素列選択部2は、仮名文字展開部1で得られた複数の音素列の中から、実際の音声サンプルに適したものを選択するための処理を行なうようになっている。従って、仮名文字展開部1で得られた音素列が1種類だけのものであった場合、音素列選択部2における処理は省略される。

【0035】音素列選択部2における実際の処理は、各音素列候補に対する認識スコア(尤度)を求めて、認識スコアの大きい順に各音素列候補に順位付けを行ない、最大の認識スコアを示した候補を音素セグメンテーション部3に与えるというものである。

【0036】音素列選択部2で使用される音声認識手法は、仮名文字展開部1で得られる音素列の候補数が、通常の単語であれば多くても数十～百程度の範囲に収まることを考えると、一般に数百語程度の認識が可能な認識手法であればどのようなものであっても構わない。

【0037】ただし、音素列選択部2での認識処理は、一般の音声認識に比べて、その識別対象がどれもかなり類似したものとなるため、候補間の僅かな差異を的確に識別することのできる高い認識能力を持つものである必要がある。

【0038】また、音素セグメンテーション部3は、音素列選択部2で選択された音素列に従って音声データの各音素の境界位置を算出するようになっている。具体的に、音素の境界位置の算出処理、すなわち、音素セグメンテーションを実行する方法としては、Viterbi探索による方法が知られている。

【0039】Viterbi探索による方法は、与えられた音響パラメータに対して、音素ラベル列に従って音素HMM(HMM:隠れマルコフモデル)を連結した単語HMMを適用し、最適状態経路を探索するというものである。これにより、最適経路に基づく音素境界を一意に決定することができる。

【0040】しかし、実際の音声は、調音結合の影響などにより、明確な音素境界を決定できない場合も多く、Viterbi探索による方法で得られる音素境界情報にもかなりの曖昧性(誤差)が含まれていることが予想される。

【0041】そこで、より有効な音素境界情報として、

音素境界位置だけでなくその信頼度を定量的に表すことのできる何らかの指標を導入したい。そこで、本発明では、N個(N>1)の音素列候補の探索経路から得られる複数の音素境界情報からそのばらつき(分散)を求め、これを信頼度の指標として用いることができる。ここで、N個(N>1)の音素列候補としては、候補の中で上位第1位から第N位までの候補が用いられる。この上位第1位から第N位までの候補を、以下では、N-best候補と称する。また、探索経路としては、Viterbi経路を用いることができる。

【0042】一般に、N-best候補(複数候補)のViterbi経路を求める場合、第2位候補以下には、第1位候補の経路とごく一部分のアライメントのみが異なるような経路が大量に湧き出してくる。この場合、個々の経路から得られる情報量が少ないため、仮にN-best候補(複数候補)の探索を行なっても、10音素程度から成る音声に対して全ての音素境界のばらつきを推定できる十分な数の候補を得るためには、候補の数(Nの値)を相当大きく(数百～数千程度に)しなければならない。このような探索を単純なViterbi探索法の応用(N-best対応化)で行なうことは非現実的である。

【0043】そこで、そのような探索を高速で実現することができ、大語彙の音声認識手法としても実績のあるA*探索法を使用することができる。A*探索法には、最適解が高速に探索できることに加えて、高速かつ柔軟なN-best解の探索が可能であるという大きな利点がある。ここでは、この特徴を活かして、候補数を1000程度とするN-best候補の算出を行なう。すなわち、音素セグメンテーション部3は、大量の候補(大量のN-best候補)を高速に算出するための手段として、A*探索法を利用し、各候補から得られたそれぞれの音素境界位置の平均値および分散を求めることで、各音素境界位置を正規分布として求めることができる。

【0044】このように、音素セグメンテーション部3は、音声データに対してHMM(隠れマルコフモデル)に代表される音響モデルを使用してN個(N>1)の音素列候補(N-best候補)の探索経路(Viterbi経路)を算出することで、仮名文字展開部1で展開された個々の音素の境界位置の点推定値(平均値)だけでなく、その区間推定値(分散)についても算出し、仮名文字展開部1で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部1で展開された個々の音素の境界位置の信頼度を表す指標として用いるようになっている。

【0045】なお、この音素境界位置の正規分布を求める際に、N個の候補から得られる音素境界位置から単純に平均値や分散を計算するのではなく、各候補から得られる音素境界位置に対して、その候補のスコアに応じた

重み付けを行なった後で平均値や分散を計算することで、最終的に得られる正規分布の信頼性の向上を図ることも可能である。

【0046】すなわち、音素セグメンテーション部3は、N個(N>1)の候補(N-best候補)から仮名文字展開部1で展開された個々の音素の境界位置の平均および分散を求める際に、N個の探索経路(Viterbi経路)から求められる音素の境界位置の情報に対して、その経路のスコアに応じた重みを乗じて集計することで、より信頼性の高い平均値および分散値を求めることができる。

【0047】音素のラベリング処理は、現在の技術レベルでは完全に自動化することは難しい。そのため、最終的には、人間が自動的に得られた結果の妥当性を判断し、必要に応じて編集を行なう必要がある。

【0048】音素境界位置エディタ部4は、仮名文字展開部1、音素列選択部2、音素セグメンテーション部3で得られた結果を作業者に分かり易く提示するようになっており、これによって、編集作業の支援を図ることが可能に構成されている。

【0049】図2は音素境界位置エディタ部4の画面表示例を示す図である。

【0050】図2において、(a)は音素列提示窓であり、音素列提示窓(a)には、仮名文字展開部1で得られた複数の音素列が、音素列選択部2で得られたスコアの順に表示されている。初期状態では、図2にハッチングで示すように最も高いスコアを持つ音素列が選択されているが、別の候補を選択することで、音素セグメンテーション部3に与える音素列を変更することができる。

【0051】また、図2において、(b)は全体波形表示窓であり、全体波形表示窓(b)は、編集作業の対象となっている音声波形全体や、音素セグメンテーション部3で得られた各音素の境界位置を表示するためのものである。なお、全体波形表示窓(b)には、後述のように、拡大波形表示窓/音素境界編集用窓(c)に表示する部分波形の範囲指定用窓(e)も併せて表示される。

【0052】また、図2において、(c)は拡大波形表示/音素境界編集用窓であり、拡大波形表示/音素境界編集用窓(c)は、全体波形表示窓(b)や、全体波形表示窓(b)内に表示される範囲指定窓(e)で選択された部分の波形を表示するための窓である。なお、拡大波形表示/音素境界編集用窓(c)に表示される部分波形の範囲指定窓(e)の大きさは、マウス操作によって自由に伸縮することができ、その結果に応じて拡大する範囲を変更することができる。また、拡大波形表示/音素境界編集用窓(c)において、音素境界位置の変更を、この窓(c)内に表示されるカーソル(d)を移動することで行なうことができるようになっている。

【0053】カーソル(d)を移動する場合には、以下のモードを選択することができる。

【0054】すなわち、第1のモードとして、音素セグメンテーション部3で得られている各音素境界の正規分布の値をそのまま表示するモードを選択できる。この第1のモードでは、音素境界位置エディタ部4は、音声セグメンテーション部3において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値を提示する(表示する)ことによって、編集中の音素境界位置の妥当性を作業者に提示することができる。

【0055】また、第2のモードとして、音素セグメンテーション部3で得られている各音素境界の正規分布の値に応じて、カーソルの色を変化させるモードを選択できる。この第2モードでは、音素境界位置エディタ部4は、音素セグメンテーション部3において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値に応じて、カーソルの色を変化させる(例えば信頼度が高い時には赤で、信頼度が低くなるに従って、赤→黄→緑→青を連続的に変化させる)ことによって、編集中の境界位置の妥当性を直感的に分かり易い形で作業者に提示することができる。

【0056】また、第3のモードとして、音素セグメンテーション部3で得られている各音素境界の分散の値に応じて、カーソルの移動可能範囲に制限を設けるモードを選択できる。この第3モードでは、音素境界位置エディタ部4は、音素セグメンテーション部3において得られたそれぞれの音素境界位置の区間推定値(分散)に応じて、マニュアル操作で移動可能な音素境界位置の範囲に制限を設けることができる。

【0057】また、上記第1、第2、第3のモードを適宜組み合わせることもできる。

【0058】なお、拡大波形表示/音素境界編集用窓(c)で行なわれた音素境界の編集結果は、直ちに全体波形表示窓(b)内に表示されている音素境界位置にも反映される。

【0059】このように、本発明では、仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に仮名文字展開部1で展開し、仮名文字展開部1で展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を音素列選択部2で選択させ、音素列選択部2で選択された音素列に従って音声データの各音素の境界位置を音素セグメンテーション部3で算出し、音素列選択部2および音素セグメンテーション部3で得られた結果を音素境界位置エディタ部4で対話的に修正するようになっているので、高精度な音素ラベル付き音声データベースを半自動的に作成することができる。

【0060】図3は図1の音声データベース作成装置のハードウェア構成例を示す図である。図3を参照すると、この音声データベース作成装置は、例えばワークステーションやパーソナルコンピュータ等で実現され、全

体を制御するCPU21と、CPU21の制御プログラム等が記憶されているROM22と、CPU21のワークエリア等として使用されるRAM23と、キーボードやマウスなどの操作部24と、ディスプレイ26とを有している。

【0061】ここで、CPU21は、図1の仮名文字展開部1、音素列選択部2、音素セグメンテーション部3、音素境界位置エディタ部4の機能を有している。

【0062】なお、CPU21におけるこのような仮名文字展開部1、音素列選択部2、音素セグメンテーション部3、音素境界位置エディタ部4等としての機能は、例えばソフトウェアパッケージ(具体的には、CD-ROM等の情報記録媒体)の形で提供することができ、このため、図3の例では、情報記録媒体30がセットさせるとき、これを駆動する媒体駆動装置31が設けられている。

【0063】換言すれば、本発明の音声データベース作成装置は、操作部、ディスプレイ等を備えた汎用の計算機システムにCD-ROM等の情報記録媒体に記録されたプログラムを読み込ませて、この汎用計算機システムのマイクロプロセッサに音声データベース作成処理を実行させる装置構成においても実施することが可能である。この場合、本発明の音声データベース作成処理を実行するためのプログラム(すなわち、ハードウェアシステムで用いられるプログラム)は、媒体に記録された状態で提供される。プログラムなどが記録される情報記録媒体としては、CD-ROMに限られるものではなく、ROM、RAM、フレキシブルディスク、メモリカード等が用いられても良い。媒体に記録されたプログラムは、ハードウェアシステムに組み込まれている記憶装置、例えばハードディスク装置にインストールされることにより、このプログラムを実行して、音声データベース作成処理機能を実現することができる。

【0064】

【発明の効果】以上に説明したように、請求項1乃至請求項9記載の発明によれば、仮名文字で記述された読みを、それを発声した場合に出現し得る音素列候補に展開する仮名文字展開部と、仮名文字展開部で展開された音素列候補の中で、実際の音声データに最も良く適合する音素列を選択する音素列選択部と、音素列選択部で選択された音素列に従って音声データの各音素の境界位置を算出する音素セグメンテーション部と、音素列選択部および音素セグメンテーション部で得られた結果を対話的に修正するための音素境界位置エディタ部とを有しているので、高精度な音素ラベル付き音声データベースを半自動的に作成することができる。すなわち、実際の音声サンプルに対応する音素列の決定や個々の音素境界位置の決定という、一般に知識や経験が要求される作業が自動化されるため、音声データベースの作成に必要な人的労力が軽減され、未熟練者でも高品質な音声データベ-

スの作成することができる。

【0065】特に、請求項2記載の発明によれば、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用いるようになっており、自動的に推定された音素境界位置に含まれる誤差の可能性に関する情報を信頼度という形でデータベースの利用者に提供することができるため、利用者側での対処が容易になる。

【0066】また、請求項3記載の発明によれば、請求項2記載の音声データベース作成装置において、前記音素セグメンテーション部は、複数の候補から仮名文字展開部で展開された個々の音素の境界位置の平均および分散を求める際に、複数の探索経路から求められる音素の境界位置の情報に対して、その経路のスコアに応じた重みを乗じて集計することで、より信頼性の高い平均値および分散値を求めることができる。すなわち、音素境界位置の算出時に、各候補のスコアを利用するために、より推定精度の高い音素境界情報を算出することが可能になる。

【0067】また、請求項4記載の発明によれば、請求項2記載の音声データベース作成装置において、前記音素セグメンテーション部は、大量の候補を高速に算出するために、A*探索法を利用ようになっており、この場合には、候補の探索において多数の候補算出が可能になるため、そこから得られる音素境界位置の点推定値(平均値)や区間推定値(分散)といった統計量の信頼度を高めることが可能になる。

【0068】また、請求項5記載の発明によれば、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音声セグメンテーション部において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値を提示するので、編集時の音素境界位置の妥当性を作業者に提示することができ(換言すれば、自動的に推定された音素境界位置の信頼度の値を表示することによって、編集作業の妥当性を作業者に正確に提示することができる)、音素境界位置の編集結果の質を高い

レベルに維持することが可能になる。

【0069】また、請求項6記載の発明によれば、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音素セグメンテーション部において得られたそれぞれの音素境界位置の信頼度を表す正規分布から求められる音素境界位置の信頼度の値に応じて、カーソルの色を変化させるので、編集集中の境界位置の妥当性を直感的に分かり易い形で作業者に提示することができ（換言すれば、自動的に推定された音素境界位置の信頼度に応じてカーソルの色を変化させることによって、編集作業の妥当性を直感的に分かり易い形で作業者に提示することができ）、音素境界位置の編集結果の質を高いレベルに維持することが可能になる。

【0070】また、請求項7記載の発明によれば、請求項1記載の音声データベース作成装置において、前記音素セグメンテーション部は、音声データに対して所定の音響モデルを使用して複数の音素列候補の探索経路を算出することで、仮名文字展開部で展開された個々の音素

の境界位置の平均値および分散を算出し、仮名文字展開部で展開された個々の音素の境界位置の平均値と分散によって定義される正規分布を、仮名文字展開部で展開された個々の音素の境界位置の信頼度を表す指標として用い、また、前記音素境界位置エディタ部は、音素セグメンテーション部において得られたそれぞれの音素境界位置の分散に応じて、マニュアル操作で移動可能な音素境界位置の範囲に制限を設けるので（すなわち、分散が小さい音素境界に関しては、作業者の未熟さに起因するミスなどによってその位置が大きく変更されることがないように強い制約を設け、分散が大きい音素境界に関してはその値に応じてある程度自由に變更可能にすることによって）、音素境界位置の編集結果の質を高いレベルに維持することが可能になる。

【図面の簡単な説明】

【図1】本発明に係る音声データベース作成装置の構成例を示す図である。

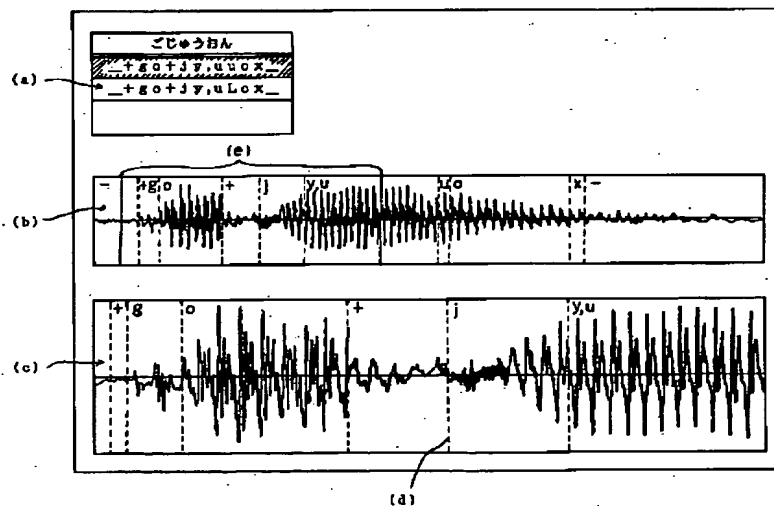
【図2】音素境界位置エディタ部の画面表示例を示す図である。

【図3】図1の音声データベース作成装置のハードウェア構成例を示す図である。

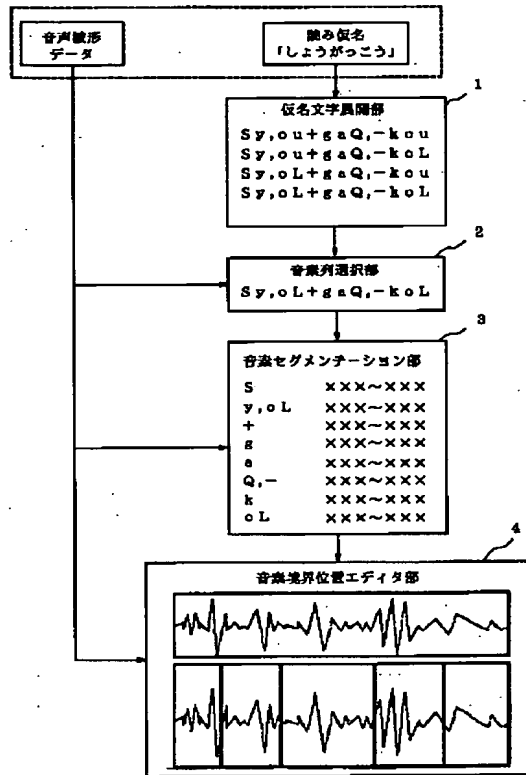
【符号の説明】

- | | |
|---|--------------|
| 1 | 仮名文字展開部 |
| 2 | 音素列選択部 |
| 3 | 音素セグメンテーション部 |
| 4 | 音素境界位置エディタ部 |

【図2】



【図1】



【図3】

